

Another generalization of Mason's ABC-theorem

Michiel de Bondt

June 24, 2009

The well-known ABC-conjecture is generally formulated as follows:

The ABC-conjecture. *Consider the set S of triples $(A, B, C) \in \mathbb{N}^3$ such that $ABC \neq 0$, $\gcd\{A, B, C\} = 1$ and*

$$A + B = C$$

Then for every $\epsilon > 0$, there exists a constant K_ϵ such that

$$C \leq K_\epsilon \cdot R(ABC)^{1+\epsilon}$$

for all triples $(A, B, C) \in S$, where $R(ABC)$ denotes the square-free part of the product ABC .

The ABC-conjecture is studied in many papers, and this article will not be another of them. Instead, we consider an analog of this conjecture for polynomials over \mathbb{C} instead of integers: Mason's ABC-theorem:

Mason's ABC-theorem. *Let f_1, f_2, f_3 be polynomials over \mathbb{C} without a common factor, not all constant, such that*

$$f_1 + f_2 + f_3 = 0$$

Then

$$\max_{1 \leq m \leq 3} \deg f_m \leq r(f_1 f_2 f_3) - 1$$

where $r(g)$ denotes the number of distinct zeros of g .

This theorem was proved at first by Stothers in [13]. So Mason did what Stayman did with the bridge convention that has his name: he made the theorem known, even popular.

The bound in Mason's theorem can be reached by examples of arbitrary large degree, namely $f_1 = f^3, f_2 = ig^2, f_3 = -(f^3 - g^2)$, where f and g reach H. Davenport's bound:

$$\deg(f^3 - g^2) \geq \frac{1}{2} \deg f + 1$$

All f and g that reach the Davenport bound are determined in [17]. The easiest example is

$$(x^2 + 2)^3 - (x^3 + 3x)^2 = 3x^2 + 8$$

So Mason's theorem seems the best you can get. But there is room for generalization. One direction is followed for the ABC-conjecture as well, namely adding more integers/polynomials to (get) the sum that vanishes. Another direction is allowing more indeterminates in the polynomials. We will discuss both generalizations. There has already been done a lot of work in these direction, mainly using so called *Wronskians*, but it seems that no one has combined all ideas to get the best generalized results one can get by means of Wronskians.

A third direction of generalization is to use elements of so-called *function fields* instead of univariate polynomials [3, 5, 16], or using *meromorphic functions* instead of multivariate polynomials [6]. These generalizations will decrease the readability of this expository paper, so we restrict ourselves to polynomials.

1 Generalizations of Mason's ABC-theorem

Let p be a (possibly multivariate) polynomial over \mathbb{C} . Then we can factorize p :

$$p = p_1^{e_1} p_2^{e_2} \cdots p_s^{e_s}$$

with all p_i irreducible and pairwise relatively prime, and all $e_i \geq 1$. Let

$$\mathfrak{r}(p) := p_1 p_2 \cdots p_s$$

be the square-free part of p and denote by $r(p)$ the degree of $\mathfrak{r}(p)$.

Associating polynomials with principal ideals, we have that $\mathfrak{r}(p)$ is the radical of p ; hence the symbol \mathfrak{r} is used.

Mason's ABC-theorem for three polynomials is generally formulated as follows [7, 11–13]:

Theorem 1.1. *Let f_1, f_2, f_3 be pairwise relatively prime univariate polynomials (in the same variable) over \mathbb{C} , not all constant, such that*

$$f_1 + f_2 + f_3 = 0$$

Then

$$\max_{1 \leq m \leq 3} \deg f_m \leq r(f_1 f_2 f_3) - 1$$

In [10, Theorem 1.2], H.N. Shapiro and G.H. Sparer generalize theorem 1.1 as follows, see also [6]:

Theorem 1.2. *Let $n \geq 3$ and f_1, f_2, \dots, f_n be pairwise relatively prime (possibly multivariate) polynomials over \mathbb{C} , not all constant, such that*

$$f_1 + f_2 + \cdots + f_n = 0$$

Then

$$\max_{1 \leq m \leq n} \deg f_m \leq (n-2) \left(r(f_1 f_2 \cdots f_n) - 1 \right)$$

In [1, Theorem 5], M. Bayat and H. Teimoori formulate the following improvement of the estimation bound of theorem 1.2 (so with all f_i 's pairwise relatively prime) as follows: they replace $(n-2)(r(f_1 f_2 \cdots f_n) - 1)$ by

$$(n-2) \left(r(f_1 f_2 \cdots f_n) - \frac{n-1}{2} \right)$$

for the case that at most one of the f_i 's is constant and by

$$(n-k-1) \left(r(f_1 f_2 \cdots f_n) - \frac{n-k}{2} \right)$$

for the case that exactly $k \geq 1$ of the f_i 's are constant. This is indeed an improvement, for if $k < n$ of the f_i 's are constant, then $n-k-1 \leq n-2$ and

$$r(f_1 f_2 \cdots f_n) \geq n-k \geq \frac{n-k}{2} \geq 1$$

because there cannot be exactly one f_i that is not constant

Unfortunately, the proof of [1, Theorem 5] is incorrect: [1, Lemma 4] has counterexamples. But we shall see that the theorem itself is correct. In [5], the univariate case of theorem 1.2 is proved, and also the erratic [1, Theorem 5] can be viewed as a correct proof for the univariate case.

But let us first discuss the condition that the f_i 's are pairwise relatively prime. This condition is quite restrictive, so it is a good idea to try and get rid of it, and replace it by something weaker. The example $n = 3$, $f_1 = f_2 = x^{100}$, $f_3 = -2x^{100}$ shows that we cannot just forget the condition that all f_i 's are relatively prime. So let us replace it by the condition that just

$$\gcd\{f_1, f_2, \dots, f_n\} = 1 \tag{1}$$

Now theorem 1.2 remains valid for $n = 3$, because the conditions $\gcd\{f_1, f_2, f_3\} = 1$ and $f_1 + f_2 + f_3 = 0$ imply that f_1, f_2, f_3 are pairwise relatively prime.

This is no longer the case if $n \geq 4$. Reading the proof of theorem 1.2 above as given in [10], it seems that $r(f_1 f_2 \cdots f_n)$ is just a shorthand notation for $r(f_1) + r(f_2) + \cdots + r(f_n)$, but if the f_i 's are not relatively prime, then both expressions are different. So we replace $r(f_1 f_2 \cdots f_n)$ by $r(f_1) + r(f_2) + \cdots + r(f_n)$ as well. There are, however, also generalizations with $r(f_1 f_2 \cdots f_n)$, which we will discuss later.

Now the example $n = 4$, $f_1 = -f_2 = x^{100}$, $f_3 = -f_4 = (x+1)^{100}$ shows us that we are not ready yet to prove something. The problem is that $f_1 + f_2 + \cdots + f_n$ has a proper subsum that vanishes. Actually, such proper subsums can be seen as instances of the original sum with smaller n , and it seems reasonable that (1) is satisfied for these subsums as well, i.e.

$$f_{i_1} + f_{i_2} + \cdots + f_{i_s} = 0 \implies \gcd\{f_{i_1}, f_{i_2}, \dots, f_{i_s}\} = 1$$

where $1 \leq i_1 < i_2 < \dots < i_s \leq n$. This way we get a valid assertion:

Theorem 1.3. *Let $n \geq 3$ and f_1, f_2, \dots, f_n be (possibly multivariate) polynomials over \mathbb{C} , not all constant, such that*

$$f_1 + f_2 + \dots + f_n = 0$$

Assume furthermore that for all $1 \leq i_1 < i_2 < \dots < i_s \leq n$,

$$f_{i_1} + f_{i_2} + \dots + f_{i_s} = 0 \implies \gcd\{f_{i_1}, f_{i_2}, \dots, f_{i_s}\} = 1$$

Then

$$\max_{1 \leq m \leq n} \deg f_m \leq (n-2) \left(r(f_1) + r(f_2) + \dots + r(f_n) - 1 \right) \quad (2)$$

If we replace the constant term -1 on the right hand side of (2) by $+n$, then the case in which the f_i 's are univariate without a vanishing proper subsum of $f_1 + f_2 + \dots + f_n$ follows from [3, Th. B] and the proof of [3, Cor. II]. An improvement of the proof of [3, Cor. II] as indicated in section 5 below subsequently replaces the term $+n$ by $+(n-1)/2$.

If one does not wish to replace $r(f_1 f_2 \dots f_n)$ by $r(f_1) + r(f_2) + \dots + r(f_n)$ (and neither requires the f_i 's to be prime by pairs), then one can use the inequality $r(f_i) \leq r(f_1 f_2 \dots f_n)$ to obtain a coefficient $n(n-2)$, but in [14] and [3, Cor. I], it is shown that in the univariate case, $(n-1)(n-2)/2$ is enough and that -1 can be maintained within the parentheses. We will prove the multivariate version of this result:

Theorem 1.4. *Under the conditions of theorem 1.3,*

$$\max_{1 \leq m \leq n} \deg f_m \leq \frac{(n-1)(n-2)}{2} \left(r(f_1 f_2 \dots f_n) - 1 \right) \quad (3)$$

2 Improvements of theorems 1.3 and 1.4

But theorems 1.3 and 1.4 are not the best one can get. One improvement on 1.4 is by U. Zannier in [16], but his idea also applies to 1.3. The coefficient $n-2$ in (2) should be expressed in the dimension d of the vector space over \mathbb{C} spanned by the f_i 's. Since $f_1 + f_2 + \dots + f_n = 0$, d is at most $n-1$, so the straightforward improvement is replacing $n-2$ by $d-1$. But also the residual term $(n-2) \cdot -1$ can be improved: the natural improvement of the corresponding term $(n-1)(n-2)/2$ in (9) of [1, Theorem 5] is $d(d-1)/2$, so we get

$$\max_{1 \leq m \leq n} \deg f_m \leq (d-1) \left(r(f_1) + r(f_2) + \dots + r(f_n) - \frac{d}{2} \right)$$

Another improvement is due to P.-C. Hu and C.-C. Yang in [5, 6]. They extend the definition of the $r(g)$ by defining

$$\mathfrak{r}_e(g) = \gcd\{g, \mathfrak{r}(g)^e\}$$

and $r_e(g) = \deg \mathfrak{r}_e(g)$. So $\mathfrak{r}_1(g) = \mathfrak{r}(g)$ is the square-free part of g and $\mathfrak{r}_2(g)$ is the cube-free part of g , etc. Now we have a trivial inequality

$$r_e(g) \leq e r(g)$$

and taking $e = n - 2$ indicates precisely how Hu and Yang improve the estimate: they migrate the coefficient $n - 2$ to a subscript of r . This migration has the drawback that the residual term $(n - 2) \cdot -1$ does not survive several reductions any more (reductions that decrease the dimension of the vector space over \mathbb{C} spanned by the f_i 's). This can be overcome by only stating that there is a ρ with $2 \leq \rho \leq n - 1$, such that

$$\max_{1 \leq m \leq n} \deg f_m \leq (\rho - 1) \left(r(f_1) + r(f_2) + \cdots + r(f_n) - \frac{\rho}{2} \right)$$

and combining the above idea with that of Zannier, we even assume that $\rho \leq d$ instead of $\rho \leq n - 1$.

Theorem 2.1. *Let $n \geq 3$ and f_1, f_2, \dots, f_n be (possibly multivariate) polynomials over \mathbb{C} , not all constant, such that*

$$f_1 + f_2 + \cdots + f_n = 0$$

Assume furthermore that for all $1 \leq i_1 < i_2 < \cdots < i_s \leq n$,

$$f_{i_1} + f_{i_2} + \cdots + f_{i_s} = 0 \implies \gcd\{f_{i_1}, f_{i_2}, \dots, f_{i_s}\} = 1$$

Now let d be the dimension of the vector space over \mathbb{C} spanned by the f_i 's. Then there exists a ρ with $2 \leq \rho \leq d$, such that

$$\max_{1 \leq m \leq n} \deg f_m \leq r_{\rho-1}(f_1) + r_{\rho-1}(f_2) + \cdots + r_{\rho-1}(f_n) - \frac{\rho(\rho-1)}{2} \quad (4)$$

$$\leq (d' - 1) \left(r(f_1) + r(f_2) + \cdots + r(f_n) - \frac{d'}{2} \right) \quad (5)$$

for all d' between d and $n - k + 1$ inclusive, where k is the number of constant f_i 's.

Proof of [1, Theorem 5]. Since $f_1 + f_2 + \cdots + f_n = 0$, it follows that $d \leq n - 1$. So the first inequality (9) of [1, Theorem 5] follows. Assume that exactly k of the f_i 's are constant for some k with $1 \leq k \leq n - 1$, and assume without loss of generality that f_n is not constant. Since the vector space over \mathbb{C} spanned by the k constant f_i 's has dimension 1 at most, the vector space over \mathbb{C} spanned by f_1, f_2, \dots, f_{n-1} has dimension $(n - 1) - (k - 1) = n - k$ at most. But since $f_1 + f_2 + \cdots + f_n = 0$, the latter vector space is also the vector space over \mathbb{C} spanned by f_1, f_2, \dots, f_n . So $d \leq n - k$ and the second inequality (10) of [1, Theorem 5] follows as well. \square

The improvements on theorem 1.4 are similar to those on theorem 1.3:

Theorem 2.2. *Under the conditions of theorem 2.1, there exists a σ with $1 \leq \sigma \leq d(d-1)/2$ such that*

$$\max_{1 \leq m \leq n} \deg f_m \leq r_\sigma(f_1 f_2 \cdots f_n) - \sigma \quad (6)$$

$$\leq \frac{d'(d'-1)}{2} (r(f_1 f_2 \cdots f_n) - 1) \quad (7)$$

for all $d' \geq d$.

We postpone the proofs of theorems 2.1 and 2.2 until section 6, since we first consider some applications.

3 Applications to Fermat-Catalan equations

Just like the ABC-conjecture for integers can be used to tackle Fermat's Theorem for integers, versions of Mason's Theorem can be used to tackle polynomial Diophantic equations:

Theorem 3.1 (Generalized Fermat-Catalan). *Assume*

$$g_1^{e_1} + g_2^{e_2} + \cdots + g_n^{e_n} = 0$$

and f_1, f_2, \dots, f_n satisfy the conditions of theorem 2.1, where $f_i = g_i^{e_i}$ for all i . Then

$$\sum_{i=1}^n \frac{1}{e_i} > \frac{1}{d-1}$$

where d is the dimension of the vector space over \mathbb{C} spanned by the f_i 's

Proof (based on ideas in [5]). Assume f_m has the largest degree among the f_i 's. From theorem 2.1, and $r(f_i) \leq \deg g_i = e_i^{-1} \deg f_m$, it follows that

$$\deg f_m \leq (d-1) \left(\sum_{i=1}^n \frac{1}{e_i} \deg f_m - \frac{d}{2} \right)$$

which rewrites to

$$\left(\sum_{i=1}^n \frac{1}{e_i} - \frac{1}{d-1} \right) \deg f_m \geq \frac{d}{2} \quad (8)$$

which completes the proof. \square

In [10, Th. 3.1] and [1, Th. 8], theorem 3.1 is proved by way of the following inequality:

$$\left(\sum_{i=1}^n \frac{1}{e_i} - \frac{1}{d-1} \right) \sum_{i=1}^n \deg g_i \geq \frac{d}{2} \sum_{i=1}^n \frac{1}{e_i} \quad (9)$$

but the proof of (9) will not be copied in a third article today.

In [10, (3.3)] and [1, Cor. 10], the result of theorem 3.1 is rewritten into a Fermat-type equation, i.e. with all e_i equal. But it is not observed that in the Fermat case, the condition that the f_i 's are relatively prime by pairs can be omitted. Having a version of a generalized Mason's theorem in which the f_i 's must be relatively prime by pairs is only partially an excuse for that, since it suffices to use the case that f_1, f_2, \dots, f_{n-1} are linearly independent of theorem 1.3, which can be proved with the methods of [10] and [1], see also [5, 6, Th. 1.3].

We say that polynomials f_1 and f_2 are *similar* if $f_2 = \lambda f_1$ for some $\lambda \in \mathbb{C}^*$.

Theorem 3.2 (Generalized Fermat). *Assume*

$$g_1^d + g_2^d + \dots + g_n^d = 0$$

for some polynomials g_i , not all zero, and suppose that

$$d \geq n(n-2)$$

Then the vanishing sum $g_1^d + g_2^d + \dots + g_n^d$ decomposes into vanishing subsums

$$g_{i_1}^d + g_{i_2}^d + \dots + g_{i_s}^d = 0$$

with $1 \leq i_1 < i_2 < \dots < i_s \leq n$, for which all g_{i_j} 's are pairwise similar.

Proof. Assume without loss of generality that $g_1 \neq 0$. Since $g_1^d + g_2^d + \dots + g_n^d = 0$, g_1^d is contained in the vector space over \mathbb{C} spanned by g_2^d, \dots, g_n^d . Assume without loss of generality that

$$g_2^d, \dots, g_l^d$$

is a basis of this vector space and that

$$g_1^d = \lambda_2 g_2^d + \dots + \lambda_s g_s^d$$

with $s \leq l \leq n$ and $\lambda_2 \dots \lambda_s \neq 0$. In order to reduce to the case that the g_i 's are relatively prime and $d = n-1$, we define

$$h_i := \frac{\sqrt[d]{\lambda_i} g_i}{\gcd\{g_1, g_2, \dots, g_s\}}$$

for all $i \leq s$, where $\lambda_1 = -1$, since then we get

$$h_1^d + \dots + h_s^d = 0$$

Furthermore, h_2^d, \dots, h_s^d are linearly independent over \mathbb{C} , and

$$\gcd\{h_1^d, \dots, h_s^d\} = \gcd\{h_1, \dots, h_s\}^d = 1$$

In order to prove this theorem, it suffices to show that all h_i 's are constant at this stage. So assume that this is not the case. Then it follows from theorem 3.1 that

$$\frac{s}{d} > \frac{1}{s-2}$$

i.e. $d < s(s-2) \leq n(n-2)$. Contradiction, so all h_i 's are constant. \square

4 A theorem of Davenport

Now let us look at sums of powers that do *not* vanish:

$$g_1^{e_1} + g_2^{e_2} + \cdots + g_{n-1}^{e_{n-1}} = g_n \neq 0$$

and suppose that no subsum of $g_1^{e_1} + g_2^{e_2} + \cdots + g_{n-1}^{e_{n-1}}$ vanishes. Now the question is how far the degree of g_n can drop. In [4], H. Davenport studied the case $n = 3$, $e_1 = 3$, $e_2 = 2$, and showed that

$$\deg(f^3 - g^2) \geq \frac{1}{2} \deg g + 1$$

see also [13]. We shall formulate a generalization of this result that improves [5, (6)], by weakening the conditions.

But first, we need some preparations. Notice that (7) of theorem 2.2 follows immediately from (6), once you realize that not all f_i 's are constant. It is somewhat more work to get (5) of theorem 2.1 from (4). At first, we remark that we can take all constant f_i 's together, resulting in exactly one constant f_i if they do not cancel out and no constant f_i 's if they do. This reduction alters k and n . But $n - k$ is not affected and d only might decrease by one, whence the range of d' is at least preserved. Next, it suffices to prove that

$$\left((d' - 1) - (\rho - 1) \right) \left(r(f_1) + r(f_2) + \cdots + r(f_n) \right) \geq \frac{d'(d' - 1)}{2} - \frac{\rho(\rho - 1)}{2}$$

which follows since the right hand side equals $\rho + (\rho + 1) + \cdots + (d' - 1) \leq ((d' - 1) - (\rho - 1))(d' - 1)$ and

$$r(f_1) + r(f_2) + \cdots + r(f_n) \geq n - k \geq d' - 1$$

where $k \leq 1$ now.

If d' is bounded by $n - k$ instead of $n - k + 1$ (and such a d' exists for $d \leq n - k$), then one of the f_i 's, say f_n , does not need to be estimated in order to boost the residual term to $d'(d' - 1)/2$:

$$\max_{1 \leq m \leq n} \deg f_m \leq (d' - 1) \left(r(f_1) + r(f_2) + \cdots + r(f_{n-1}) - \frac{d'}{2} \right) + r_{\rho-1}(f_n)$$

Estimating $r_{\rho-1}(f_n)$ by $\deg f_n$ and realizing that at least two f_i 's have maximum degree, we get (10) of theorem 4.1 below under the conditions of theorem 2.1:

Theorem 4.1. *Let f_1, f_2, \dots, f_n be (possibly multivariate) polynomials over \mathbb{C} , not all similar, such that*

$$f_1 + f_2 + \cdots + f_n = 0$$

Assume furthermore that for all $1 \leq i_1 < i_2 < \cdots < i_s \leq n$,

$$f_{i_1} + f_{i_2} + \cdots + f_{i_s} = 0 \implies \deg \gcd\{f_{i_1}, f_{i_2}, \dots, f_{i_s}\} \leq \deg f_n$$

Let d be the dimension of the vector space over \mathbb{C} spanned by the f_i 's. Then

$$\max_{1 \leq m \leq n-1} \deg f_m - \deg f_n \leq (d' - 1) \left(r(f_1) + r(f_2) + \cdots + r(f_{n-1}) - \frac{d'}{2} \right) \quad (10)$$

for all d' between d and $n - k$ inclusive, where k is the number of constant f_i 's. Furthermore, equality is only possible in (10) if $\gcd\{f_1, f_2, \dots, f_n\} = 1$.

Proof. Take $m' \leq n - 1$ such that $\max_{1 \leq m \leq n-1} \deg f_m = \deg f_{m'}$. We reduce to the case that the conditions of theorem 2.1 are satisfied. If f_n is constant, then the conditions of theorem 2.1 are satisfied and hence we are done. So assume that f_n is not constant. Then we can remove all constant f_i 's and add them to f_n without affecting the estimate, because $\deg f_n$ and $n - k$ will not change due to this maneuver. Furthermore, subsums $f_{i_1} + f_{i_2} + \cdots + f_{i_s} = 0$ for which $\gcd\{f_{i_1}, f_{i_2}, \dots, f_{i_s}\} \neq 1$ are not affected. Now we distinguish two cases.

- There is a minimal vanishing subsum of $f_1 + f_2 + \cdots + f_n = 0$ that contains both $f_{m'}$ and f_n as summands.
Assume without loss of generality that $f_{m'} + f_{m'+1} + \cdots + f_n = 0$ and let $h := \gcd\{f_{m'}, f_{m'+1}, \dots, f_n\}$. Then

$$\begin{aligned} & \deg f_{m'} - \deg f_n \\ &= \deg \frac{f_{m'}}{h} - \deg \frac{f_n}{h} \\ &\leq (d' - 1) \left(r\left(\frac{f_{m'}}{h}\right) + r\left(\frac{f_{m'+1}}{h}\right) + \cdots + r\left(\frac{f_{n-1}}{h}\right) - \frac{d'}{2} \right) \\ &\leq (d' - 1) \left(r(f_{m'}) + r(f_{m'+1}) + \cdots + r(f_{n-1}) - \frac{d'}{2} \right) \\ &\leq (d' - 1) \left(r(f_1) + r(f_2) + \cdots + r(f_{n-1}) - \frac{d' + m' - 1}{2} \right) \end{aligned}$$

where d' is at least the dimension of the vector space spanned by $f_{m'}, f_{m'+1}, \dots, f_n$ and at most $n - m' + 1$, and equality is only possible if h is constant and $m' = 1$. This gives the desired result.

- There is no minimal vanishing subsum of $f_1 + f_2 + \cdots + f_n = 0$ that contains both $f_{m'}$ and f_n as summands.
Assume without loss of generality that $f_1 + f_2 + \cdots + f_{m'} = 0$ and let $h := \gcd\{f_1, f_2, \dots, f_{m'}\}$. Then $\deg h \leq \deg f_n$. In case $f_1, f_2, \dots, f_{m'}$ are all similar, then the left hand side of (10) is zero and the right hand side is positive, as desired. So assume that that is not the case. By (5) in theorem 2.1,

$$\begin{aligned} & \deg f_{m'} - \deg f_n \\ &\leq \deg \frac{f_{m'}}{h} \\ &\leq (d' - 1) \left(r\left(\frac{f_1}{h}\right) + r\left(\frac{f_2}{h}\right) + \cdots + r\left(\frac{f_{m'}}{h}\right) - \frac{d'}{2} \right) \end{aligned}$$

$$\begin{aligned}
&\leq (d' - 1) \left(r(f_1) + r(f_2) + \cdots + r(f_{m'}) - \frac{d'}{2} \right) \\
&\leq (d' - 1) \left(r(f_1) + r(f_2) + \cdots + r(f_{n-1}) - \frac{d' + n - m' - 1}{2} \right)
\end{aligned}$$

where d' is at least the dimension of the vector space spanned by $f_1, f_2, \dots, f_{m'}$ and at most $m' + 1$, and equality is not possible because $m' = n - 1$ implies $f_n = 0$. This gives the desired result. \square

Now substitute $f_i = g_i^{e_i}$ for all $i \leq n - 1$ and also $f_n = -g_n = \sum_{i=1}^n g_i^{e_i}$, in (10). Then

$$\left(\sum_{i=1}^{n-1} \frac{1}{e_i} - \frac{1}{d' - 1} \right) \max_{1 \leq m \leq n-1} \deg g_m^{e_m} \geq \frac{d'}{2} - \frac{1}{d' - 1} \deg \sum_{i=1}^{n-1} g_i^{e_i} \quad (11)$$

follows from (10) in a similar way as (8) follows from (5) of theorem 2.1, see also [5, (6)].

Indeed, applying (11) on the sum $f^3 + (ig)^2$ gives $-\frac{1}{6} \deg(f^3) \geq 1 - \deg(f^3 - g^2)$ for $d' = 2$, which is equivalent to $\deg(f^3 - g^2) \geq \frac{1}{2} \deg f + 1$. For $d' = 3$, we get $\frac{1}{3} \deg f^3 \geq \frac{3}{2} - \frac{1}{2} \deg(f^3 - g^2)$, i.e. $\deg(f^3 - g^2) \geq 3 - 2 \deg f$, which is useless.

By replacing n by $n + 1$ in (11), we obtain the following from theorem 4.1.

Theorem 4.2. *Assume*

$$g_1^{e_1} + g_2^{e_2} + \cdots + g_n^{e_n} \neq 0$$

and no subsum of $g_1^{e_1} + g_2^{e_2} + \cdots + g_n^{e_n}$ vanishes. Then

$$\left(\sum_{i=1}^n \frac{1}{e_i} - \frac{1}{d' - 1} \right) \max_{1 \leq m \leq n} \deg g_m^{e_m} \geq \frac{d'}{2} - \frac{1}{d' - 1} \deg \sum_{i=1}^n g_i^{e_i}$$

for all d' between d and $n + 1$ inclusive, where d is the dimension of the vector space over \mathbb{C} spanned by $g_1^{e_1}, g_2^{e_2}, \dots, g_n^{e_n}$. Furthermore, equality cannot be reached in case $\gcd\{g_1, g_2, \dots, g_n\} \neq 1$.

In [17], it is proved that for all even degrees of f , there are univariate polynomials f, g over \mathbb{C} such that $\deg(f^3 - g^2) = \frac{1}{2} \deg f + 1$. Now assume $\deg(f^3 - g^2) = \frac{1}{2} \deg f + 1$. Then $\gcd\{f, g\} = 1$ and the Mason bound on $-f^3 + g^2 + (f^3 - g^2) = 0$ gives us

$$\deg f^3 \leq r_1(fg(f^3 - g^2)) - 1 \leq \deg(fg(f^3 - g^2)) - 1$$

which is bound to be an equality. Furthermore, $fg(f^3 - g^2)$ is bound to be square-free. But any linear combination $\lambda f^3 + \mu g^2$ with $\lambda\mu \neq 0$ is bound to be square-free, since otherwise the inequality

$$\deg f^3 \leq \frac{1}{2} \left(r_1(f^3) + r_1(g^2) + r_1(f^3 - g^2) + r_1(\lambda f^3 + \mu g^2) - 1 \right)$$

would be violated. The above estimate is an instance of (12) in section 5 below, since there exists a vanishing linear combination without zero coefficients of the arguments of r_1 on the right hand side.

5 Some discussion on theorems 2.1 and 2.2

We describe now why the condition that all f_i 's are relatively prime by pairs is needed in [1, 5, 6, 10]. They reduce to the case of maximal dimension $d = n - 1$ as follows. Assume that f_n has the largest degree and say that f_1, f_2, \dots, f_d is a basis of the vector space over \mathbb{C} spanned by f_1, f_2, \dots, f_n . Then

$$f_n = \lambda_1 f_1 + \lambda_2 f_2 + \dots + \lambda_d f_d$$

for some $\lambda_i \in \mathbb{C}$. The greatest common divisor of the f_i 's in the above sum is still the same as in the original sum, but some f_i 's might have a coefficient λ_i that is zero; say that $\lambda_1 \lambda_2 \dots \lambda_\rho \neq 0$ and $\lambda_{\rho+1} = \lambda_{\rho+2} = \dots = \lambda_d = 0$. Then

$$\lambda_1 f_1 + \lambda_2 f_2 + \dots + \lambda_\rho f_\rho + (-f_n) = 0$$

is a vanishing sum of maximal dimension ρ . But the problem is that the greatest common divisor of the f_i 's in the last sum might be larger than that of the original sum.

But the above method does work when each set of d f_i 's generates the whole vector space over \mathbb{C} spanned by the f_i 's, because that implies that $\rho = d$ above. So in this case one can get the estimates of theorems 2.1 and 2.2. But one can get even better estimates in this particular case, namely

$$\max_{1 \leq m \leq n} \deg f_m \leq \frac{1}{n-d} \left(r_{\rho-1}(f_1) + r_{\rho-1}(f_2) + \dots + r_{\rho-1}(f_n) - \frac{\rho(\rho-1)}{2} \right) \quad (12)$$

and

$$\max_{1 \leq m \leq n} \deg f_m \leq \frac{1}{n-d} (r_\sigma(f_1 f_2 \dots f_n) - \sigma) \quad (13)$$

combining techniques of [6] and the proof of [16, Th. 2], and also ideas in section 7 to get $\rho(\rho-1)/2 \leq \sigma$. We sketch the proof at the very end of this article.

In [2, Th. 2] it is shown that the coefficient $d'(d'-1)/2$ of (7) in theorem 2.2 cannot be replaced by something less than $2n-5$, and the author conjectures that this coefficient can indeed be improved to $2n-5$, i.e.

$$\max_{1 \leq m \leq n} \deg f_m \leq (2n-5)(r(f_1 f_2 \dots f_n) - 1)$$

I did not find similar considerations on (5) in theorem 2.1 in literature. So let us do something ourselves. The factor $(d'-1)$ in (5) cannot be improved, as is shown by the example

$$\begin{aligned} f_i &= \binom{n-2}{i-1} (x^{10^{100}})^{i-1} \quad (1 \leq i < n) \\ f_n &= -(x^{10^{100}} + 1)^{n-2} \end{aligned}$$

The term $d'/2$ in (5) cannot be improved to $3d'/4$, as is shown by the example

$$\begin{aligned} f_i &= \lceil n/2 \rceil \binom{\lceil n/2 \rceil (\lfloor n/2 \rfloor + 1) - 2}{\lceil n/2 \rceil i - 1} x^{\lceil n/2 \rceil i - 1} \quad (i \leq \lfloor n/2 \rfloor) \\ f_i &= -\zeta_{\lceil n/2 \rceil}^i \left(x + \zeta_{\lceil n/2 \rceil}^i \right)^{\lceil n/2 \rceil (\lfloor n/2 \rfloor + 1) - 2} \quad (i > \lfloor n/2 \rfloor) \end{aligned}$$

for the case that none of the f_i 's is constant, and by the example

$$\begin{aligned} f_i &= \lceil n/2 \rceil \binom{\lceil n/2 \rceil \lfloor n/2 \rfloor - 1}{\lceil n/2 \rceil (i - 1)} x^{\lceil n/2 \rceil (i - 1)} \quad (i \leq \lfloor n/2 \rfloor) \\ f_i &= -\zeta_{\lceil n/2 \rceil}^{-i} \left(x + \zeta_{\lceil n/2 \rceil}^i \right)^{\lceil n/2 \rceil \lfloor n/2 \rfloor - 1} \quad (i > \lfloor n/2 \rfloor) \end{aligned}$$

for the case that f_1 is constant, but it might be possible to improve it to $3(d' - 1)/4$.

In section 4, we have reduced (5) in theorem 2.1 to (4) and (7) in theorem 2.2 to (6). Therefore it remains to prove (4) and (6). But before we do that, we ask ourselves the question whether (4) and (6) can be seen as instances of one single, more general estimate. [3] has some valuable ideas in that direction. Under the extra assumption that the f_i 's are univariate and $d = n - 1$, (7) for $d' = d = n$ follows immediately from [2, Cor. I], and [2, Cor. II] implies

$$\max_{1 \leq m \leq n} \deg f_m \leq (n - 2)(r(f_1) + r(f_2) + \cdots + r(f_n) + 1)$$

but, since the f_i 's are linearly independent, the number k of constant f_i 's is at most 1. Since the number of empty S_i 's in [2, Cor. II] equals k as well, one can improve [2, Cor. II] to

$$\begin{aligned} H(u_1, u_2, \dots, u_n) &\leq (n - 2) \left(|S_1| + |S_2| + \cdots + |S_n| + k - \frac{n + 1}{2} \right) - \\ &\quad \frac{(n - 1)(n - 2)}{2} (2g - 2) \end{aligned} \quad (14)$$

and (5) in theorem 2.1 for $d' = d = n$ follows.

The proof of (14) is left as an exercise to the interested reader. The general result that implies both [2, Col. I] and (the improved version (14) of) [2, Col. II] is [2, Theorem A].

The rest of this article is organized as follows. In sections 6 to 8, we prove (4) of theorem 2.1 and (6) of theorem 2.2. In section 6, we reduce to the univariate case. In section 7, we present the Wronskian, the key element in all generalized versions of Mason's theorem, except [14]. Section 8 consists of the actual proofs of (4) and (6). At last, in section 9, we combine (4) and (6) with ideas of [2].

6 Some reductions of the main theorem

By replacing the original sum by the minimal vanishing subsum containing $f_{m'}$ as a term, where $\deg f_{m'} = \max_{1 \leq m \leq n} \deg f_m$, we see that in order to prove (4)

of theorem 2.1 and (6) of theorem 2.2, we can restrict ourselves to the case that $f_1 + f_2 + \dots + f_n$ has no proper subsum that vanishes.

We show now that we can restrict ourselves to the case that the f_i 's are univariate. More particular, a generic substitution $x_i = p_i y + q_i$ will do the reduction. Assume that no proper subsum of $f_1 + f_2 + \dots + f_n$ vanishes and say that there are l variables in the f_i 's. Let G be the set of nonempty proper subsums

$$f_{i_1} + f_{i_2} + \dots + f_{i_s}$$

and

$$\bar{G} = \{\bar{g} \mid g \in G\}$$

where \bar{g} is the largest degree homogeneous part of g (i.e. the sum of all terms that have the same degree as g). Now pick a $p \in \mathbb{C}^l$ such that

$$\bar{g}(p) \neq 0$$

for all $\bar{g} \in \bar{G}$ (a p that has coordinates that are transcendental over the field of coefficients of the \bar{g} 's will do).

Assume without loss of generality that $p_1 \neq 0$ and define

$$\hat{f}_i := f_i(p_1 x_1, x_2 + p_2 x_1, \dots, x_l + p_l x_1)$$

for all i . Since $\gcd\{f_1, f_2, \dots, f_n\} = 1$, $\gcd\{\hat{f}_1, \hat{f}_2, \dots, \hat{f}_n\} = 1$ as well. So if we apply the extended gcd-theorem with respect to x_1 , we find $a_i \in \mathbb{C}(x_2, \dots, x_l)[x_1]$ such that

$$1 = a_1 \hat{f}_1 + a_2 \hat{f}_2 + \dots + a_n \hat{f}_n$$

For each i , write $a_i = \sum_{j=1}^{\infty} a_{i,j} x_1^j$ with $a_{i,j} \in \mathbb{C}(x_2, \dots, x_l)$ and only finitely many $a_{i,j}$ nonzero. Now put $q_1 := 0$ and take $(q_2, \dots, q_l) \in \mathbb{C}^{k-1}$ such that the denominators of the nonzero $a_{i,j}$'s do not vanish on (q_2, \dots, q_l) . Then

$$\begin{aligned} 1 &= a_1(q_2, \dots, q_l)[x_1] \hat{f}_1(x_1, q_2, \dots, q_l) + \\ &\quad a_2(q_2, \dots, q_l)[x_1] \hat{f}_2(x_1, q_2, \dots, q_l) + \dots + \\ &\quad a_n(q_2, \dots, q_l)[x_1] \hat{f}_n(x_1, q_2, \dots, q_l) \end{aligned} \tag{15}$$

Put

$$\tilde{f}_i := \hat{f}_i(y, q_2, \dots, q_l) = f_i(q + yp) = f_i(p_1 y + q_1, p_2 y + q_2, \dots, p_l y + q_l)$$

for all i . From (15), it follows that $\gcd\{\tilde{f}_1, \tilde{f}_2, \dots, \tilde{f}_n\} = 1$.

Since $r_{\rho-1}(\tilde{f}_i) \leq r_{\rho-1}(f_i)$ for all i and $r_{\sigma}(\tilde{f}_1 \tilde{f}_2 \dots \tilde{f}_n) \leq r_{\sigma}(f_1 f_2 \dots f_n)$, it suffices to show that $\deg \tilde{f}_i = \deg f_i$ for all i and no proper subsum of $\tilde{f}_1 + \tilde{f}_2 + \dots + \tilde{f}_n = 0$ vanishes. We do so by proving that for all proper subsets I of $\{1, 2, \dots, n\}$:

$$\deg \left(\sum_{i \in I} \tilde{f}_i \right) = \deg \left(\sum_{i \in I} f_i \right)$$

i.e.

$$\deg g(q + yp) = \deg g$$

for all $g \in G$. This is true, since the coefficient of $y^{\deg g}$ in $g(q + yp)$ is equal to $\bar{g}(p)$, which is nonzero by assumption.

7 The Wronskian

Let f_1, f_2, \dots, f_n be polynomials in one and the same variable, say y . Then the *Wronskian determinant* of f_1, f_2, \dots, f_n is defined as

$$W(f_1, f_2, \dots, f_n) := \det \begin{pmatrix} f_1 & f_2 & \cdots & f_n \\ f_1' & f_2' & \cdots & f_n' \\ \vdots & \vdots & \ddots & \vdots \\ f_1^{(n-1)} & f_2^{(n-1)} & \cdots & f_n^{(n-1)} \end{pmatrix}$$

and the *Wronskian matrix* is the corresponding matrix on the right hand side.

Since differentiating is a linear operator, it follows that $W(f_1, f_2, \dots, f_n) = 0$ in case

$$\lambda_1 f_1 + \lambda_2 f_2 + \cdots + \lambda_n f_n = 0 \quad (16)$$

for some nonzero $\lambda \in \mathbb{C}^n$. Now a classical theorem tells us that the reverse is true as well: if f_1, f_2, \dots, f_n are linearly independent (i.e. (16) implies $\lambda = 0$), then $W(f_1, f_2, \dots, f_n) \neq 0$. The example $f_1(x) = x^3$, $f_2(x) = |x|^3$ shows us that the f_i 's need to be polynomials.

Despite that the oldest known proof of this theorem by Frobenius is elementary, we give another proof, inspired by the proof of [15, Lm. 8]. The reason for that will be given below.

So let us assume that f_1, f_2, \dots, f_n are linearly dependent. If there are two f_i 's with the same degree, then we can subtract a multiple of the first from the second to reduce the degree of the second, since this operation does not affect the Wronskian determinant. Progressing in this direction gives us that all f_i 's have different degrees. Now order the f_i 's by increasing degrees. This might only change the sign of the Wronskian determinant.

The matrix

$$\begin{pmatrix} f_1^{(\deg f_1)} & f_2^{(\deg f_1)} & \cdots & f_n^{(\deg f_1)} \\ f_1^{(\deg f_2)} & f_2^{(\deg f_2)} & \cdots & f_n^{(\deg f_2)} \\ \vdots & \vdots & \ddots & \vdots \\ f_1^{(\deg f_n)} & f_2^{(\deg f_n)} & \cdots & f_n^{(\deg f_n)} \end{pmatrix}$$

is upper triangular and does not have zeros on the diagonal. Hence, its deter-

minant does not vanish. Since it is a submatrix of

$$M := \begin{pmatrix} f_1 & f_2 & \cdots & f_n \\ f'_1 & f'_2 & \cdots & f'_n \\ f_1^{(2)} & f_2^{(2)} & \cdots & f_n^{(2)} \\ \vdots & \vdots & \ddots & \vdots \\ f_1^{(\deg f_n)} & f_2^{(\deg f_n)} & \cdots & f_n^{(\deg f_n)} \end{pmatrix}$$

this latter matrix has full rank n . Now we can make a square matrix M' of full rank n out of M by throwing away redundant rows of M , i.e. throwing away rows that are dependent of the rows above it. It suffices to prove that M' is the Wronskian matrix, i.e.

$$M' = \begin{pmatrix} f_1 & f_2 & \cdots & f_n \\ f'_1 & f'_2 & \cdots & f'_n \\ f_1^{(2)} & f_2^{(2)} & \cdots & f_n^{(2)} \\ \vdots & \vdots & \ddots & \vdots \\ f_1^{(n-1)} & f_2^{(n-1)} & \cdots & f_n^{(n-1)} \end{pmatrix}$$

Write $f^{(i)}$ for the vector

$$(f_1^{(i)}, f_2^{(i)}, \dots, f_n^{(i)})$$

and $f = f^{(0)}$ and $f' = f^{(1)}$. Assume that the m -th row of M' is $(f^{(m-1)})^t$, but the $(m+1)$ -th row of M' is not $(f^{(m)})^t$, say it is $(f^{(j)})^t$ with $j > m$. Then $(f^{(j-1)})^t$ is in the space generated by the first m rows of M' , i.e.

$$f^{(j-1)} = a_0 f + a_1 f' + a_2 f^{(2)} + \cdots + a_{m-1} f^{(m-1)} \quad (17)$$

where the a_i are rational functions, i.e. quotients of polynomials, for all i . Differentiating (17) gives

$$f^{(j)} = (a'_0 f + a_0 f') + (a'_1 f' + a_1 f'') + \cdots + (a'_{m-1} f^{(m-1)} + a_{m-1} f^{(m)})$$

Since each of the $2m$ terms on the right hand side is contained in the space generated by the first m rows of M' , $f^{(j)}$ is contained in this space as well. Contradiction, so the m -th row of M' is $(f^{(m-1)})^t$ for all m .

In [9, Lemma 6, pp. 15-16], a generalization of the Wronskian theorem for more variables is formulated. The operators $\frac{\partial^i}{\partial y^i}$ are in fact replaced by operators Δ_i , each of which is a product of partial derivatives. The number of partial derivatives that Δ_i decomposes into, multiple appearances counted by their frequency, is called the *order* $o(\Delta_i)$ of Δ_i .

The usual Wronskian determinant is replaced by

$$W_\Delta(f_1, f_2, \dots, f_n) := \det \begin{pmatrix} \Delta_1 f_1 & \Delta_1 f_2 & \cdots & \Delta_1 f_n \\ \Delta_2 f_1 & \Delta_2 f_2 & \cdots & \Delta_2 f_n \\ \vdots & \vdots & \ddots & \vdots \\ \Delta_n f_1 & \Delta_n f_2 & \cdots & \Delta_n f_n \end{pmatrix} \quad (18)$$

and the author T. Schneider of [9] proves that if f_1, f_2, \dots, f_n are linearly independent, then $W_\Delta(f_1, f_2, \dots, f_n) \neq 0$ for certain operators Δ_i of order $i - 1$ at most. In particular, Δ_1 is the identity operator, and the first row looks the same as in the case of one variable.

Unlike the above proof of the classical Wronskian theorem, the proof of this theorem by Frobenius cannot be generalized to more indeterminates. The way Schneider proves his multivariate result is by reducing to the univariate Wronskian theorem. But his theorem does not show that there are Δ_i 's of all orders $0, 1, 2, \dots, \rho$, where ρ is the maximum order of the Δ_i 's, unlike a straightforward generalization of the above proof of the classical Wronskian theorem to more indeterminates. Neither does his methods give tools to prove that

$$W_\Delta(hf_1, hf_2, \dots, hf_n) = h^n W_\Delta(f_1, f_2, \dots, f_n) \quad (19)$$

(19) can be found in [6, Lm. 2.1]. But this lemma is somewhat different to both our methods and [9, Lemma 6, pp. 15-16], since the Wronskian determinant might be zero.

Take for instance $f = (1, xy, x^2y^2)$. Notice that

$$W_{1, \frac{\partial}{\partial x}, \frac{\partial^2}{\partial x^2}}(1, xy, x^2y^2) = \det \begin{pmatrix} 1 & xy & x^2y^2 \\ 0 & y & 2xy^2 \\ 0 & 0 & 2y^2 \end{pmatrix} = 2y^3$$

and this is also a generalized Wronskian one can get by the multivariate variant of the above method, since $\frac{\partial}{\partial y} x^i y^i = i x^i y^{i-1} = \frac{x \partial}{y \partial y} x^i y^i$. The above Wronskian matrix is however *not* of the form of [6, Lm. 2.1] and [15, Lm. 8], because $\frac{\partial}{\partial y} f$ is *not* linearly dependent over \mathbb{C} of its rows. The Wronskian matrix of both lemma's must be that of

$$W_{1, \frac{\partial}{\partial x}, \frac{\partial}{\partial y}}(1, xy, x^2y^2) = 0$$

instead.

In the proofs of theorems 2.1 and 2.2, we shall employ a special generalized Wronskian, one without an identity operator:

Lemma 7.1. *Let f_1, f_2, \dots, f_n be polynomials over \mathbb{C} in the variables y, z_1, z_2, \dots, z_l , such that each f_i is of the following form:*

$$f_i = (\lambda_{1,i} z_1 + \lambda_{2,i} z_2 + \dots + \lambda_{l,i} z_l) \cdot \tilde{f}_i$$

where \tilde{f}_i is a polynomial over \mathbb{C} in the variable y . Assume that f_1, f_2, \dots, f_n are linearly independent. Then there exists a $\Delta = (\Delta_1, \Delta_2, \dots, \Delta_n)$ with

$$W_\Delta(f_1, f_2, \dots, f_n) \neq 0$$

such that for each i , either

$$\Delta_i = \frac{\partial}{\partial z_j}$$

for some j , or (if $i \geq 2$)

$$\Delta_i = \frac{\partial}{\partial y} \Delta_{i-1}$$

Proof. Choose j such that $\lambda_{j,n} \neq 0$. Say that $\lambda_{j,1} = \dots = \lambda_{j,m} = 0$ and $\lambda_{j,m+1} \dots \lambda_{j,n} \neq 0$. We distinguish three cases:

- $\frac{\partial}{\partial z_j} f_{m+1}, \dots, \frac{\partial}{\partial z_j} f_n$ are linearly dependent.
Say that

$$\frac{\partial}{\partial z_j} f_{m+1} = \mu_{m+2} \frac{\partial}{\partial z_j} f_{m+2} + \dots + \mu_n \frac{\partial}{\partial z_j} f_n$$

Replace f_{m+1} by $f_{m+1} - (\mu_{m+2} f_{m+2} + \dots + \mu_n f_n)$ and apply induction on $-m$.

- $\frac{\partial}{\partial z_j} f_{m+1}, \dots, \frac{\partial}{\partial z_j} f_n$ are linearly independent and $m = 0$.
Then the result follows by applying the Wronskian theorem (in one variable) on $\frac{\partial}{\partial z_j} f_1, \frac{\partial}{\partial z_j} f_2, \dots, \frac{\partial}{\partial z_j} f_n$. The operators are $\Delta_i = \frac{\partial^i}{\partial y^{i-1} \partial z_j}$.
- $\frac{\partial}{\partial z_j} f_{m+1}, \dots, \frac{\partial}{\partial z_j} f_n$ are linearly independent and $m \geq 1$.
From the above case, it follows that $W_D(f_{m+1}, \dots, f_n) \neq 0$, where $D_i := \frac{\partial^i}{\partial y^{i-1} \partial z_j}$. By induction on n , we have $W_\Delta(f_1, f_2, \dots, f_m) \neq 0$. Now extend Δ by defining $\Delta_{m+i} = D_i$ for all $i \geq 1$. Since $\frac{\partial}{\partial z_j} f_i = 0$ for all $i \leq m$, it follows that

$$W_\Delta(f_1, f_2, \dots, f_n) = W_\Delta(f_1, \dots, f_m) \cdot W_D(f_{m+1}, \dots, f_n) \neq 0$$

and Δ remains of the desired form. \square

Notice that the above lemma can be generalized to more variables as well.

8 Proof of the main theorem

From the reductions in sections 4 and 6, it follows that in order to prove theorems 2.1 and 2.2, it suffices to prove the following:

Theorem 8.1. *Let $\tilde{f}_1, \tilde{f}_2, \dots, \tilde{f}_n$ be nonzero polynomials over \mathbb{C} in the variable y such that $\gcd\{\tilde{f}_1, \tilde{f}_2, \dots, \tilde{f}_n\} = 1$ and*

$$\tilde{f}_1 + \tilde{f}_2 + \dots + \tilde{f}_n = 0$$

Let d be the dimension of the vector space over \mathbb{C} spanned by the \tilde{f}_i 's and assume furthermore that no proper subsum of $\tilde{f}_1 + \tilde{f}_2 + \dots + \tilde{f}_n$ vanishes. Then

$$\max_{1 \leq m \leq n} \deg \tilde{f}_m \leq r_{\rho-1}(\tilde{f}_1) + r_{\rho-1}(\tilde{f}_2) + \dots + r_{\rho-1}(\tilde{f}_n) - \frac{\rho(\rho-1)}{2}$$

for some ρ with $2 \leq \rho \leq d$, and

$$\max_{1 \leq m \leq n} \deg \tilde{f}_m \leq r_\sigma(\tilde{f}_1 \tilde{f}_2 \cdots \tilde{f}_n) - \sigma$$

for some σ with $1 \leq \sigma \leq d(d-1)/2$.

Assume without loss of generality that $\tilde{f}_1, \tilde{f}_2, \dots, \tilde{f}_d$ is a basis of the vector space over \mathbb{C} spanned by the \tilde{f}_i 's. For each $j > d$, there exists unique $\lambda_{j,i}$ such that

$$\tilde{f}_j = \sum_{i=1}^d \lambda_{j,i} \tilde{f}_i \quad (20)$$

In order to get rid of all linear relations between the \tilde{f}_i 's except the sum relation, we define

$$f_i := \left(\sum_{j=d+1}^n \lambda_{j,i} z_j \right) \cdot \tilde{f}_i$$

for all $i \leq d$, and

$$f_i := -z_i \cdot \tilde{f}_i$$

for all $i > d$. It follows from (20) that

$$\begin{aligned} \sum_{i=1}^n f_i &= \sum_{i=1}^d \sum_{j=d+1}^n \lambda_{j,i} z_j \tilde{f}_i - \sum_{j=d+1}^n z_j \tilde{f}_j \\ &= \sum_{j=d+1}^n z_j \left(\sum_{i=1}^d \lambda_{j,i} \tilde{f}_i - \tilde{f}_j \right) \\ &= 0 \end{aligned}$$

Furthermore, it follows from (20) that

$$\sum_{i=1}^d \left(1 + \sum_{j=d+1}^n \lambda_{j,i} \right) \tilde{f}_i = \sum_{i=1}^d \tilde{f}_i + \sum_{j=d+1}^n \sum_{i=1}^d \lambda_{j,i} \tilde{f}_i = \sum_{i=1}^n \tilde{f}_i = 0$$

whence

$$\sum_{j=d+1}^n \lambda_{j,i} = -1 \quad (1 \leq i \leq d) \quad (21)$$

for $\tilde{f}_1, \tilde{f}_2, \dots, \tilde{f}_d$ are linearly independent.

Lemma 8.2. $\mu_1 f_1 + \mu_2 f_2 + \cdots + \mu_n f_n = 0$ implies $\mu_1 = \mu_2 = \cdots = \mu_n$.

Proof. Let G be the graph with vertices $\{1, 2, \dots, n\}$ and connect two vertices j, i by an edge if $\lambda_{j,i} \neq 0$. Notice that G is a bipartite graph between $\{1, 2, \dots, d\}$ and $\{d+1, \dots, n\}$. We first show that G is connected. Assume the opposite.

Say that G does not have an edge between $\{1, \dots, d', d+1, \dots, n'\}$ and $\{d'+1, \dots, d, n'+1, \dots, n\}$, where either $d' < d$ or $n' < n$. Then $\lambda_{j,i} = 0$ for all $j > n'$ and $i \leq d'$, whence by (21)

$$\sum_{j=d+1}^{n'} \lambda_{j,i} = -1 \quad (22)$$

for all $i \leq d'$. On the other hand, $\lambda_{j,i} = 0$ for all $j \leq n'$ and $i > d'$, whence

$$\sum_{j=d+1}^{n'} \lambda_{j,i} = 0 \quad (23)$$

for all $i > d'$.

Substituting $z_j = 1$ for all $j \leq n'$ and $z_j = 0$ for all $j > n'$ in $\sum_{i=1}^n f_i$, it follows from (22) and (23) that we obtain

$$\sum_{i=1}^d \left(\sum_{j=d+1}^{n'} \lambda_{j,i} \right) \tilde{f}_i - \sum_{j=d+1}^{n'} \tilde{f}_j = - \sum_{i=1}^{d'} \tilde{f}_i - \sum_{j=d+1}^{n'} \tilde{f}_j$$

which is zero, since $\sum_{i=1}^n f_i$ is zero. Since no proper subsum of $\sum_{i=1}^n \tilde{f}_i$ vanishes, we have $d' = d$ and $n' = n$. Contradiction, so G is connected.

Now assume $\mu_1 f_1 + \mu_2 f_2 + \dots + \mu_n f_n = 0$. Pick a $j > d$. Substituting $z_j = 1$ and $z_m = 0$ for all $m \neq j$ in $\sum_{i=1}^n \mu_i f_i$ gives us

$$\sum_{i=1}^d \mu_i \lambda_{j,i} \tilde{f}_i - \mu_j \tilde{f}_j = 0$$

but on account of (20), also

$$\sum_{i=1}^d \mu_j \lambda_{j,i} \tilde{f}_i - \mu_j \tilde{f}_j = 0$$

so by subtraction

$$\sum_{i=1}^d (\mu_i - \mu_j) \lambda_{j,i} \tilde{f}_i = 0$$

Since $\tilde{f}_1, \tilde{f}_2, \dots, \tilde{f}_d$ are linearly independent over \mathbb{C} , $(\mu_i - \mu_j) \lambda_{j,i} = 0$ for all $i \leq d$. So

$$\lambda_{j,i} \neq 0 \implies \mu_i = \mu_j \quad (24)$$

Since G is connected, the desired result follows. \square

From lemma 8.2, it follows that f_1, f_2, \dots, f_{n-1} are linearly independent, whence we can apply lemma 7.1 to get

$$W_{\Delta}(f_1, f_2, \dots, f_{n-1}) \neq 0$$

where $\Delta = (\Delta_1, \Delta_2, \dots, \Delta_{n-1})$ satisfies the properties of lemma 7.1. Since $f_1 + f_2 + \dots + f_n = 0$, we have

$$\begin{aligned} W_\Delta(f_1, f_2, \dots, f_{n-1}) &= (-1)^{n-i} W_\Delta(f_1, \dots, f_{i-1}, f_{i+1}, \dots, f_n) \\ &= (-1)^{n-1} W_\Delta(f_2, \dots, f_{n-1}, f_n) \end{aligned} \quad (25)$$

Let ρ be the maximum among the orders $o(\Delta_1), o(\Delta_2), \dots, o(\Delta_{n-1})$, i.e. the maximum number of partial derivatives which any Δ_m may decomposes into. Put

$$\sigma := \sum_{i=1}^{n-1} (o(\Delta_i) - 1)$$

Let $j > d$. Since $\frac{\partial}{\partial z_m} f_j = 0$ for all $j \neq m$, and the left hand side of (25) does not vanish, $\frac{\partial}{\partial z_j} \in \{\Delta_1, \Delta_2, \dots, \Delta_{n-1}\}$. A similar argument on the right hand side of (25) gives $\frac{\partial}{\partial z_n} \in \{\Delta_1, \Delta_2, \dots, \Delta_{n-1}\}$. So $n - d$ of the $n - 1$ Δ_i 's have order 1. It follows from lemma 7.1 that

$$2 \leq \rho \leq d \quad \text{and} \quad 1 \leq \frac{\rho(\rho-1)}{2} \leq \sigma \leq \frac{d(d-1)}{2}$$

Lemma 8.3.

$$\tilde{f}_1 \tilde{f}_2 \cdots \tilde{f}_n \left| \mathfrak{r}_{\rho-1}(\tilde{f}_1) \mathfrak{r}_{\rho-1}(\tilde{f}_2) \cdots \mathfrak{r}_{\rho-1}(\tilde{f}_n) \cdot W_\Delta(f_1, f_2, \dots, f_{n-1}) \right.$$

and

$$\tilde{f}_1 \tilde{f}_2 \cdots \tilde{f}_n \left| \mathfrak{r}_\sigma(\tilde{f}_1 \tilde{f}_2 \cdots \tilde{f}_n) \cdot W_\Delta(f_1, f_2, \dots, f_{n-1}) \right.$$

Proof. It suffices to prove that irreducible polynomials g over \mathbb{C} in the variable y divide the right hand side at least as often as the left hand side. So let $g \in \mathbb{C}[y]$ be irreducible. Since $\gcd\{\tilde{f}_1, \tilde{f}_2, \dots, \tilde{f}_n\} = 1$, one of the \tilde{f}_i 's is not divisible by g , say that $g \nmid \tilde{f}_1$. It follows from (25) that it suffices to show that g divides $\tilde{f}_2 \cdots \tilde{f}_n$ at most as often as

$$\mathfrak{r}_{\rho-1}(\tilde{f}_1) \mathfrak{r}_{\rho-1}(\tilde{f}_2) \cdots \mathfrak{r}_{\rho-1}(\tilde{f}_n) \cdot W_\Delta(f_2, \dots, f_{n-1}, f_n)$$

and

$$\mathfrak{r}_\sigma(\tilde{f}_1 \tilde{f}_2 \cdots \tilde{f}_n) \cdot W_\Delta(f_2, \dots, f_{n-1}, f_n)$$

Now pick any term of the determinant expression $W_\Delta(f_2, \dots, f_{n-1}, f_n)$. After permuting f_2, \dots, f_n , the term at hand becomes

$$\Delta_1 f_2 \cdot \Delta_2 f_3 \cdots \Delta_{n-1} f_n$$

Now if g divides \tilde{f}_i exactly l times and hence also f_i exactly l times, then g divides $\Delta_{i-1} f_i$ at least $l - \rho$ times, since partial derivatives kill at most one instance of a factor g in their argument. But one of the partial derivatives is a $\frac{\partial}{\partial z_j}$ which does not kill any instance of g , so g divides $\Delta_{i-1} f_i$ at least $l - (\rho - 1)$ times.

The factor $\mathfrak{r}(\tilde{f}_i)^{\rho-1}$ compensates the decrease of $\rho - 1$ factors g , so g divides $\mathfrak{r}(\tilde{f}_i)^{\rho-1}\Delta_{i-1}\tilde{f}_i$ at least as often as it divides \tilde{f}_i , and the first inequality of this lemma follows. The second inequality follows from the fact that the Δ_i 's together have σ partial derivatives of the form $\frac{\partial}{\partial y}$ that might kill instances of g . \square

Lemma 8.4.

$$\deg W_\Delta(f_1, f_2, \dots, f_{n-1}) \leq \deg(\tilde{f}_1\tilde{f}_2 \cdots \tilde{f}_{n-1}) - \sigma$$

Proof. The idea is that a partial derivative decreases the degree by one. Consider a term on the left hand side of the above formula. After reordering the f_i 's, this term becomes

$$\Delta_1 f_1 \cdot \Delta_2 f_2 \cdots \Delta_{n-1} f_{n-1}$$

Since $o(\Delta_i) \geq 1$ for all i , the degree of this term is at most $\deg(f_1 f_2 \cdots f_{n-1}) - (n-1) = \deg(\tilde{f}_1 \tilde{f}_2 \cdots \tilde{f}_{n-1})$. But there are also Δ_i 's of orders larger than one, which are responsible for the term σ . \square

Proof of theorem 8.1. Assume without loss of generality that \tilde{f}_n has the largest degree among the \tilde{f}_i 's. From lemmas 8.3 and 8.4, it follows that

$$\sum_{i=1}^n \deg \tilde{f}_i \leq r_{\rho-1}(\tilde{f}_1) + r_{\rho-1}(\tilde{f}_2) + \cdots + r_{\rho-1}(\tilde{f}_n) + \deg(\tilde{f}_1 \tilde{f}_2 \cdots \tilde{f}_{n-1}) - \sigma$$

whence

$$\deg \tilde{f}_n \leq r_{\rho-1}(\tilde{f}_1) + r_{\rho-1}(\tilde{f}_2) + \cdots + r_{\rho-1}(\tilde{f}_n) - \frac{\rho(\rho-1)}{2}$$

which is the first inequality of theorem 8.1. The second inequality follows similarly. \square

9 Joining theorems 2.1 and 2.2

The general result that implies both [2, Col. I] and (the improved version (14) of) [2, Col. II] is [2, Theorem A], which we will describe now for the polynomial case. For irreducible polynomials p , let m_p denote the number of f_i 's that is *not* divisible by p . Then [2, Theorem A] implies

$$\max_{1 \leq m \leq n} \deg f_m \leq -\binom{n-1}{2} + \sum_p \left(\binom{n-1}{2} - \binom{m_p-1}{2} \right) \quad (26)$$

where \sum_p ranges over all irreducible polynomials p . It follows from (26) that

$$\max_{1 \leq m \leq n} \deg f_m \leq -\binom{n-1}{2} + \sum_{p \nmid f_1 \cdots f_n} \left(\binom{n-1}{2} - \binom{n-1}{2} \right) + \sum_{p \mid f_1 \cdots f_n} \binom{n-1}{2}$$

which is exactly the case $d' = n - 1$ of the univariate case of (7) in theorem 2.2. In order to get a similar result on (26) and (5) in theorem 2.1, we first need some preparations. Assume

$$f_i \nmid f_{i+1} \quad (27)$$

The reason for (27) is that there exists an irreducible p that divides f_i more times than it divides f_{i+1} , say that p divides f_i $l + j$ times and f_{i+1} l times. Now replace f_i by $f_i p^j$ and f_{i+1} by $f_{i+1} p^{-j}$. Then (27) might still be the case, but the divisibility by p is not the reason any more. Furthermore, for any power q of an irreducible polynomial, q divides as many f_i 's as before. If we proceed in this direction, we finally arrive at

Proposition 9.1. *There exist h_1, h_2, \dots, h_n such that*

1. $h_1 \mid h_2 \mid \dots \mid h_n$,
2. *For any power q of an irreducible polynomial, q divides as many h_i 's as it divides f_i 's.*

Notice that $h_1 = \gcd\{f_1, f_2, \dots, f_n\} = 1$. More generally, h_i is the greatest common divisor over all subsets $\{j_1, j_2, \dots, j_i\}$ of $\{1, 2, \dots, n\}$ of $\text{lcm}\{f_{j_1}, f_{j_2}, \dots, f_{j_i}\}$.

Since m_p is also the number of h_i 's that is not divisible by p ,

$$\binom{m_p - 1}{2} = \sum_{i=2}^{m_p} (i - 2) = \sum_{\substack{2 \leq i \leq n \\ p \nmid h_i}} (i - 2)$$

whence

$$\binom{n - 1}{2} - \binom{m_p - 1}{2} = \sum_{i=m_p+1}^n (i - 2) = \sum_{\substack{1 \leq i \leq n \\ p \mid h_i}} (i - 2)$$

Summing this over all p , it follows from (26) that

$$\max_{1 \leq m \leq n} \deg f_m \leq \sum_{i=1}^n (i - 2) r(h_i) - \binom{n - 1}{2} \quad (28)$$

which implies the case $d' = n - 1$ of the univariate case of (5) in theorem 2.1, for

$$\sum_{i=1}^n r(h_i) = \sum_{i=1}^n r(f_i)$$

By $r(h_1) = 0$ and $r(h_i) \leq r(h_n)$, the case $d' = n - 1$ of the univariate case of (7) in theorem 2.2 follows from (28) as well. (28) can be improved to

$$\max_{1 \leq m \leq n} \deg f_m \leq \sum_{i=1}^n r_{i-2}(h_i) - \binom{n - 1}{2}$$

which implies (4) in theorem 2.1 for $\rho = n - 2$ and (6) in theorem 2.2 for $\sigma = (n - 1)(n - 2)/2$, since $r_i(a)r_j(b) \leq r_{i+j}(ab)$. The general multivariate result that includes both theorems 2.1 and 2.2 is

$$\max_{1 \leq m \leq n} \deg f_m \leq \sum_{i=2}^n r_{(o_{i-1})-1}(h_i) - \sigma$$

where

$$o_1 \leq o_2 \leq \cdots \leq o_{n-1}$$

are the orders of the Δ_i 's. The proof is left as an exercise to the reader.

At last we sketch the proof of (12) and (13). Assume that each set of d f_i 's forms a basis of the space generated by all f_i and order the f_i 's by increasing degree. As indicated in section 5, we do not need to multiply the f_i 's by linear forms in order to get rid of unwanted linear dependences. Similar to (25), one can prove that all sequences of d f_i 's have the same Wronskian determinant $W_\Delta(f_1, f_2, \dots, f_d)$ up to a nonzero constant in \mathbb{C} . Since each set of d f_i 's generates the whole space, the greatest common divisor of such a set is 1, whence there can only be $d - 1$ f_i 's at most that are divisible by a given irreducible polynomial p . So $h_1 = h_2 = \cdots = h_{n-d+1} = 1$ and

$$\begin{array}{l|l} f_1 f_2 \cdots f_n & h_{n-d+2} h_{n-d+3} \cdots h_n \\ & \mathfrak{r}_1(h_{n-d+2}) \mathfrak{r}_2(h_{n-d+3}) \cdots \mathfrak{r}_{d-1}(h_n) W_\Delta(f_1, f_2, \dots, f_d) \end{array} \quad (29)$$

because focusing on one irreducible divisor p , one can replace f_1, f_2, \dots, f_d on the right hand side of (29) by the d f_i 's of maximum divisibility by p . Next, since each set of d f_i 's has a polynomial of maximum degree, the $n - d$ f_i 's on the left hand side of (29) that are *not* on the right side of (29) have maximum degree. That gives the factor $1/(n - d)$ in (12) and (13).

References

- [1] M. Bayat and H. Teimoori, A new bound for an extension of Mason's theorem for functions in several variables, Arch. Math. 82 (2004), 230-239.
- [2] J. Browkin and J. Brzeziński, Some remarks on the abc-conjecture, Math. Comp. 62 (1994), no. 206, 931-939.
- [3] W.D. Brownawell and D.W. Masser, Vanishing sums in function fields, Math. Proc. Cambridge Philos. Soc. 100 (1986), 427-434.
- [4] H. Davenport, On $f^3(t) - g^2(t)$, Norske Vid. Selsk. Forh. (Trondheim) 38 (1965), 86-87.
- [5] P.-C. Hu and C.-C. Yang, Notes on a generalized abc-conjecture over function fields, Ann. Math. Blaise Pascal 8 (2001), no. 1, 61-71.

- [6] P.-C. Hu and C.-C. Yang, A note on a the abc-conjecture, *Comm. Pure Appl. Math.* LV (2002), 1089-1103.
- [7] R.C. Mason, Diophantic equations over function fields, *Londen Math. Soc. Lecture Note Series* 96, Cambridge 1984.
- [8] S. Lang, *Undergraduate Algebra*, Second Edition, Springer 1990.
- [9] T. Schneider, *Einführung in die Tranzsendenten Zahlen*, Berlin 1957.
- [10] H.N. Shapiro and G.H. Sparer, Extensions of a theorem of Mason, *Comm. Pure. Appl. Math.* XLVII (1994), 711-718.
- [11] J.H. Silverman, The S-unit equation over function fields, *Math. Proc. Cambridge Philos. Soc.* 95(2) (1984), 3-4.
- [12] N. Snyder, An alternate proof of Mason's theorem. *Elemente Math.* 55(3) (2000), 93-94.
- [13] W.W. Stothers, Polynomial identities and Hauptmoduln, *Quart. J. Math.* 32(2) (1981), 349-370.
- [14] J.F. Voloch, Diagonal equations over function fields, *Bol. Soc. Brasil. Math.* 16 (1985), 29-39.
- [15] Z. Ye, On Nevanlinna's second main theorem in projective space, *Invent. Math.* 122(3) (1995), 475-507.
- [16] U. Zannier, Some remarks on the S-unit equation in function fields, *Acta Arith.* 64 (1993), 87-98.
- [17] U. Zannier, On Davenport's bound for the degree of $f^3 - g^2$ and Riemann's Existence Theorem, *Acta Arith.* 71(2) (1995), 107-137.